

A METHOD AND SYSTEM FOR STEREO ECHO CANCELLATION FOR VOIP
COMMUNICATION SYSTEMS

CROSS-REFERENCE TO RELATED APPLICATION

This application claims priority to U.S. Provisional Patent Application No. 60/189,713, filed March 15, 2000 the contents of which is hereby incorporated by reference.

FIELD OF THE INVENTION

The present invention relates generally to telecommunications systems and more particularly to an apparatus and method for cancelling line echos in a telecommunications system.

BACKGROUND

Speech quality has become a highly competitive factor in marketing telephony systems. Line or electrical echo, a phenomenon typically caused by imperfect impedance matching of network transmission sections, may significantly degrade the overall speech quality of telephony systems.

Similarly, acoustic echo may also degrade the quality of speech in a telephony system. Acoustic echo may be seen, for example, in communication devices having a near end microphone exposed to a loudspeaker driven by a far end signal or a secondary audio signal. In a full-duplex system, simultaneous two-way communication is enabled where the local user can speak and listen to received speech simultaneously with the remote user. Such simultaneous conversation, however, may create acoustic feedback problems which occur when the near end microphone picks up the far end loudspeaker signal and directs the far end loudspeaker signal back to the remote end. As a result, the remote party may hear a delayed version of their own speech referred to as an acoustic echo.

1

5

10

15

20

25

30

35

Echo cancellers have been used to remove far end electrical and acoustic echo. Typically, echo cancellers utilize adaptive filters that model the electro/acoustical echo path. The algorithm coefficients of the filter are continuously adapted to represent the impulse response of the acoustic echo path, such as for example, between the loudspeaker and microphone or the impulse response between the transmit channel and the receive channel of the network interface. The modeled responses are then subtracted from an outgoing communication signal to yield an echo reduced communication signal. However, near end speech may act as an unwanted noise signal causing the adaptive filter to diverge. Therefore, echo cancellers typically include double talk detection logic that halts filter adaptation when near end speech is active.

Acoustic echo cancellers and electrical echo cancellers operating in the presence of a secondary audio tone, such as for example, a pulse metering tone, typically utilize multiple adaptive filters. In the case of electrical echo, separate adaptive filters are typically used to separately cancel the primary and secondary audio signals. Similarly, acoustic echo cancellers typically utilize an adaptive filter to estimate the impulse response between the microphone and loudspeaker. A second electrical echo canceller is typically implemented across the transmit and receive channels to cancel the electric reflection of signals generated by an impedance mismatch at the hybrid interface.

However, the utilization of multiple adaptive filters increases the complexity and computational intensity of the echo canceller system. In addition, double talk detection logic typically declares near end speech active whenever a local near end signal is present in either adaptive filter. Therefore, echo canceller performance may be seriously degraded in applications where a secondary audio tone or signal is present at substantially all times.

1

5

Therefore it would be advantageous to have a system and method that enables filter adaptation, and thus echo cancellation in the presence of a known secondary audio signal such as a music signal, a pulse metering tone or the sound of a computer game.

SUMMARY OF THE INVENTION

10

In one aspect of the invention, an echo canceller includes an adaptive filter having filter coefficients that may be adapted to cancel an echo of a composite far end signal having a secondary audio signal and a primary telephony signal in a near end signal.

15

In another aspect of the present invention, an echo canceller includes an adaptive filter having filter coefficients that may be adapted to cancel an echo of a composite signal comprising a secondary audio signal and a far end telephony signal in a near end signal.

20

It is understood that other embodiments of the present invention will become readily apparent to those skilled in the art from the following detailed description, wherein it is shown and described only embodiments of the invention by way of illustration of the best modes contemplated for carrying out the invention. As it will be realized, the invention is capable of other and different embodiments and the details are capable of modification in various other respects, all without departing from the spirit and scope of the present invention. Accordingly, the drawings and detailed description are to be illustrative in nature and not restrictive.

25

30

35

1

DESCRIPTION OF THE DRAWINGS

5

These and other features, aspects, and advantages of the present invention will become better understood with regard to the following description, appended claims, and accompanying drawings wherein:

FIG. 1 is a simplified schematic diagram of the two-way telephony application;

10

FIG. 2 shows the addition of a pulse metering tone in a public telephone;

FIG. 3 shows the addition of an audio/music signal in an integrated telephony-ready cable-set-top-box;

15

FIG. 4 shows the design of a VoIP system with a second adaptive filter for the removal of the echo associated with the pulse metering tone;

20

FIG. 5 is a simplified block diagram of an echo canceller having a single adaptive filter designed to adapt in the presence of a secondary audio tone in accordance with an exemplary embodiment of the present invention;

25

FIG. 6 shows the input data path wherein multiple inputs are combined to form a single reference signal input into the adaptive filter of FIG. 6, in accordance with an exemplary embodiment of the present invention;

FIG. 7 is a simplified block diagram of an echo canceller in an integrated telephony cable set-top-box/cable modem in accordance with an exemplary embodiment of the present invention;

30

FIG. 8 illustrates the parallel transfer functions modeled by the adaptive filter of FIG. 7 in accordance with an exemplary embodiment of the present invention; and

FIG. 9 shows the input data path wherein multiple inputs are combined to form a single reference signal input into the adaptive filter of FIG. 7, in accordance with an exemplary embodiment of the present invention.

35

1
DETAILED DESCRIPTION

5 An exemplary embodiment of the present invention provides an echo canceller system that is particularly well suited for use in communications networks where an echo is liable to occur in the presence of a secondary audio signal or an interference pulse metering tone. A described exemplary embodiment includes an adaptive filter that adapts in the presence of a known secondary audio signal such as for example, a music signal, a pulse metering tone or the sound of a computer game. For example, an exemplary echo canceller provides echo cancellation in a two-way telephony system while a movie is playing in the background via a cable-set-top box. Furthermore, echo cancellation may be provided during a double-talk condition in an interactive gaming environment where the double-talk condition may be caused by the sound of a computer game or an interference pulse metering tone at 12 kHz or 16 kHz.

15
20 FIG. 1 is a simplified block diagram of a two-way telephony application where an echo is generated from an impedance mismatch created by a two-wire to four-wire interface in hybrid circuitry 100. Ideally the hybrid gain would be zero to eliminate coupling between the far end and near end transmission paths. However, the operation of the hybrid 100 typically results in a frequency dependent gain of less than about one-half. Hybrid gain commonly results in line echos, such as the reflection of the talker's voice at the mismatch created by the two-four wire conversion.

25
30 A reference signal 104 is input into an adaptive filter 102 that models the transfer characteristics of the dialed telephone line (i.e. the hybrid plus the tail circuit that is the telephone line from DAC to ADC). The unwanted echo may then be canceled by subtracting filtered reference signal 102(a) from the near end digital input signal 106 in a difference operator 108. The filter adaptation is dependent on the subtracted signal 108(a) output by difference operator 108 (an error feedback).
35 Therefore, near end speech and/or noise present on the near end

1 digital input signal 106, typically referred to as the double
talk condition, may cause the adaptive filter 102 to diverge.
Thus, there should be no adaptation during the double-talk
5 condition. That is, filter adaptation should be carried out only
during those intervals when the far-end speech is present, but
near-end speech is absent.

FIG. 2 shows a simplified schematic diagram of the two-way
telephony application of FIG. 1 having a pulse metering tone
10 generator 120 that generates a pulse metering tone at a frequency
of either 12 kHz or 16 kHz during substantially all intervals.
The metering pulses allow a user to monitor the cost of a call
as it progresses. In this and other embodiments, common
reference numerals may be used to represent like components.

15 The rate at which the pulse metering tone generator 120
transmits the metering pulses varies depending on the rate per
minute of the call. The minimum gap between pulses is, by way
of example, on the order of about 100 msec but may be much
greater for inexpensive local calls. The amplitude of the
20 metered pulses can be adjustable to accommodate impedance changes
during the on hook, off hook and ringing states. The harmonics
associated with this pulse metering tone within the 4 kHz
bandwidth may provide false feedback information, that may cause
the adaptive filter 102 to diverge, consequently degrading the
25 performance of the echo canceller.

FIG. 3 shows a simplified schematic diagram of the two-way
telephony application of FIG. 1 wherein an audio signal or a
music signal 130 is present on the far end signal resulting in
near end acoustic feedback. The audio signal may result for
30 example, from an integrated telephony-ready, cable-set-top-box.
The music signal may be the result of a movie or a video game 132
playing out a speaker 134 simultaneously with a two-way phone-
call, using speaker 133 and a microphone 136. Thus, the music
signal is acting either as a noise source or as the near-end
35 talker's voice. This additional signal may create a double-talk

1 condition at substantially all times. Therefore, the echo
canceller should disable adaptation since the music signal may
cause the filter 140 to diverge.

5 FIG. 4 shows a simplified schematic diagram of a two-way
VoIP system including multiple adaptive filters 150 and 152. The
second adaptive filter 152 operates at a sampling rate of 48 kHz.
The higher sampling rate is necessary to avoid aliasing since the
pulse metering tone has a frequency of 12kHz or 16 kHz. The echo
10 signal associated with this pulse metering tone may be canceled
by subtracting filtered metered tone 152(a) from near end digital
input signal 154 in a difference operator 156. However, the
utilization of the second adaptive filter 152, at an increased
sampling rate, generally increases the complexity and
15 computational intensity of the echo canceller system.

FIGS. 5 and 6 show simplified schematic diagrams of an
exemplary stereo echo cancellation system that provides echo
cancellation in the presence of secondary audio tones, such as
for example, pulse metering tones. Referring to FIG. 5, the
20 described exemplary echo cancellation system combines a
downsampled pulse metering tone 236(a) from a decimator 237 and
a far end reference signal 202(a) into a single 8 kHz combined
reference signal 237(a) through a combiner 237. The combined
reference signal 237(a) may then be input into a single 8 kHz
25 adaptive filter 200 rather than using two adaptive filters, one
at 8 kHz and the other at 48 kHz as described in connection with
FIG. 4. Therefore, the described exemplary embodiment requires
less memory and processing resources.

30 In the described exemplary embodiment of FIGS. 5 and 6, the
adaptive filter 200 may be a linear transversal filter or other
suitable finite impulse response filter. The adaptive filter 200
models the transfer characteristics of the dialed telephone line
circuit. The unwanted echo is canceled by subtracting filtered
reference signal 200(a) from near end digital input signal 204
35 in a difference operator 206. Near end speech and/or noise

1

5

10

15

20

25

30

35

present on the near end signal 204, typically referred to as the double talk condition, may cause the adaptive filter 200 to diverge. Therefore, double talk logic 212 detects near end speech and disables adaptation of the adaptive filter 200 when near end speech is active. One of skill in the art will appreciate that the present invention is not limited to a particular algorithm for detecting the double talk condition. However, conventional approaches that utilize energy estimates introduce delay into the detection of near end speech by the energy estimators (typically low pass filters) that may significantly corrupt the feedback error signal used for filter adaptation. Therefore, the described exemplary embodiment of FIGS. 5 and 6 utilize a sample based double talk detection algorithm that provides rapid detection of near end speech in the presence of far end speech. Look ahead capability may also be included so that the double talk logic may halt filter adaptation (i.e. freeze the filter taps or coefficients) before the near end speech samples reach the difference operator 206. A downsampler 208 decimates a 96 kHz near end Tx data 210 at a ratio of 12:1. The downsampler 208 forwards the decimated near end Tx data samples 208(a) to the double talk logic at an 8 kHz sample rate.

To minimize delay, the downsampler 208 does not low pass filter the near end Tx data samples 210 prior to decimation. Aliasing components that may be created are insignificant because the output of the downsampler 208(a) simply drives the double talk detection logic 212 and is not transmitted to the far end. An energy estimator 214 estimates the background noise level of the decimated near end Tx data signal 208(a) and forwards the estimated level to the double talk logic 212. The energy estimator 212 may be a low pass filter with a long time constant, on the order of about 10 seconds. With a long time constant the energy estimator tends to track the minimum energy level of the decimated near end Tx data signal 208(a). Similarly, a second

1

energy estimator 217 estimates the short term energy of the combined reference signal 237(a).

5

To avoid divergence, the double talk logic 212 selectively enables the adaptive filter 200. The double talk logic 212 utilizes a sample based algorithm to detect the presence of near end speech without incurring the delays associated with conventional systems in accordance with the following equation:

10

$$| \text{Near} | > H * \text{Peak}(| \text{Ref}_{in}(n) |) + \text{Background_Noise}(n)$$

15

The double talk logic 212 declares near end speech active when the absolute value of the near end Tx data signal 210 is greater than the product of the hybrid gain (H) and a peak statistic of the combined reference signal 237(a) summed with the estimated background noise of the decimated near end Tx data signal 210. The hybrid gain is generally a constant less than about one-half. The background noise for a typical voice channel is on the order of about -70 dBm which is far less than average active speech levels, which are typically on the order of about -25 dBm. The background noise estimate is therefore initialized to a value of about -70 dBm and thereafter periodically updated by the energy estimator 214. The peak statistic of the combined reference signal is defined by the following logic:

20

25

If $\max\{A * [| \text{Ref}_{in}(n) |, \dots, | \text{Ref}_{in}(n-L-1) |]\} > \text{Peak}(n-1)$
then

$$\text{Peak}(n) = \max \{A [| \text{Ref}_{in}(n) |, \dots, | \text{Ref}_{in}(n-L-1) |]\}$$

else

$$\text{Peak}(n) = d * \text{Peak}(n-1);$$

30

where A is a weighting function that is greater than zero and less than or equal to one. The parameter L is the number of samples over which the maximum is determined, typically in the range of zero to one hundred and twenty eight samples. In the described exemplary embodiment L is on the order of about sixty
four samples. The parameter d is a constant that is also greater

35

1

than zero and less than or equal to one and on the order of about 0.99 in the described exemplary embodiment.

5

10

Therefore, to determine the peak statistic of the far end, the double talk logic applies a weighting factor A to the absolute value of the current sample ($\text{Ref}_{in}(n)$) and previous L samples ($\text{Ref}_{in}(n-L)$). If the maximum product is greater than the previous peak statistic $\text{Peak}(n-1)$ then the current peak statistic $\text{Peak}(n)$ is set at the maximum of the product of the weighting factor and far end samples. Otherwise the current peak statistic $\text{Peak}(n)$ is set equal to d times the value of the previous peak statistic $\text{Peak}(n-1)$.

15

20

In the described exemplary embodiment, A , L and d are empirically determined to optimize performance and computational load for a given application. For example, double logic 212 can more accurately detect near end speech if the maximum is determined over a larger number of samples L . However, computational intensity also increases with increasing number of samples L . A and d can be inversely related to the number of sample L , so that A and d are smaller for larger number of samples and vice versa.

25

30

35

In the described exemplary embodiment, there is a delay associated with decimator 220. However, downsampler 208 does not low pass filter the near end TX data 210 and therefore does not have a corresponding delay. Therefore, the double talk logic 200 receives downsampled near end Tx data samples 208(a) with negligible delay and can process near end Tx data samples prior to their arrival at the difference operator 206. Thus, the delay associated with the second decimator 220 provides a look-ahead of M samples allowing the double talk logic 212 to disable adaptation of the adaptive filter 200, M samples before the near-end signal reaches the difference operator 206. The look ahead capability M is equivalent to the delay associated with the second decimator 220 and is typically two to three 8kHz samples for a ITU-T G712 compliant system.

1

In the described exemplary embodiment, an interpolater 230 up-samples the signal from the far end (Rin) from 8 to 48 kHz. The interpolator 230 may be implemented with a FIR filter that may be either minimum phase or linear phase. A minimum phase filter is preferred for voice applications that require low group delay but may tolerate group delay distortion that may be introduced by the minimum phase filter.

5

10

In addition, a metering pulse generator 232 may be used to generate metering pulses 232(a) at a frequency of 12/16 kHz.

15

In the described exemplary embodiment, a decimator 236 decimates or down samples the metering pulses 232(a) so that the sample rate of the music signal matches the sample rate of the far end reference signal 202(a), i.e. 8 kHz in the described exemplary embodiment. The decimated pulse metering tones 236(a) are forwarded to the combiner 237. The combiner 237 combines the decimated pulse metering tones 236(a) and the far end reference signal 202(a) into a single 8 kHz input 237(a) for the adaptive filter 200. The metering pulses are summed with the interpolated signal 230(a) by summer 234.

20

25

A second interpolater 238 upsamples the summed signal output by the summer 234 from 48 to 96 kHz. The interpolator 238 may again be implemented with a FIR filter. A digital to analog converter 240 converts the digital interpolated signal to an analog signal which is then transmitted to the near end. Reflections at the two - four wire conversion at the hybrid 166 may create an echo including metering tone echos and far end speech echo that is fed back to the near end.

30

An analog to digital converter 242 receives analog near end voice samples and converts the received analog signal to a digital format. The decimator 220 downsamples the received digital near end Tx data signal 210 from 96 kHz to an 8 kHz rate. The decimator 220 can be implemented with a FIR filter. The decimated signal 204 along with the output of the adaptive filter 200(a) are input to the difference operator 206 that cancels

35

1 unwanted echo by subtracting the decimated signal 204 from the filter output signal 200(a).

5 Referring to FIG. 6, the combiner 237 of FIG. 5 may include a first buffer 250 that receives the far end reference signal 202(a) and a second buffer 252 that receives the decimated pulse metering tone 236(a). Buffer outputs 250(a) and 252(a) are then combined in a third buffer 254 whose output 254(a) is forwarded to the adaptive filter 200. One of skill in the art will appreciate that a single buffer could be used to directly combine the decimated metering tones 236(a) and far end reference signal 202(a).

10 The adaptive filter 200 may be based upon a normalized least mean square algorithm (NLMS) as described in S. Haykin, Adaptive Filter Theory, and T. Parsons, Voice and Speech Processing, the contents of which are incorporated herein by reference as if set forth in full. Referring back to FIG. 5, an error signal 206(a) at the output of the difference operator 206 that is feedback to the adaptive filter 200 for filter adaptation may be characterized as follows:

$$e(n) = Tx(n) - \sum_{j=0}^{M+N+1} w_{11}(j) \text{Ref}_{in}(n-j) \text{ Where } j=0,1,2 \dots M+N+1$$

15 where $e(n)$ is the error signal 206(a) at time n , $\text{Ref}_{in}(n)$ is the combined reference signal 237(a) at time n , $Tx(n)$ is the decimated Tx data signal 204 input into the difference operator 206 at time n , and $w_{11}(j)$ are the coefficients of the transversal filter W_{11} .

20 In the described exemplary embodiment, data inputs from the first data buffer 250 are summed from $j = 0, 1, \dots M$ and data inputs from the second data buffer 252 are summed from $j = M+1, M+2, \dots M+N+1$. Thus, $\text{Ref}_{in}(0)$ is the most recent input 250(a) from the first data buffer 250 and $\text{Ref}_{in}(M)$ is the oldest input from the first data buffer. Therefore, $\text{Ref}_{in}(M-j)$ is the output of buffer one 250 delayed by $M-j$ samples. Similarly, $\text{Ref}_{in}(M+1)$

1 is the most recent input 252(a) from the second data buffer 252 and $\text{Ref}_{in}(M+N+1)$ is the oldest input from the second data buffer 252.

5 In the described exemplary embodiment, M and N are parameters that may be selected to match the electrical path. Therefore in one embodiment, the dimension of the transversal filter may be the worst case echo path length. In the described exemplary embodiment, filter coefficients $W_{11}(j)$, for $j=0$ to 10 $M+N+1$, are given by:

$$W_{11}(j) = w_{11}(j) + \mu * e(n) * \text{Re } f_{in}(n-j)$$

15 wherein $w_{11}(j)$ is preferably initialized to a reasonable value such as for example zero.

Assuming a block size of four msec (or 32 samples at a sampling rate of 8 kHz), the short term average energy of the reference signal E_{ref} is the sum of the last 32 reference samples so that the convergence gain may be given by:

$$\mu = \frac{\alpha}{E_{ref}(n)}$$

20 where α is the adaptation step size and E_{ref} is the sum of the squares of those samples $\text{Ref}_{in}(n)$ that are saved in the third 25 buffer 254. In the described exemplary embodiment, α is set to zero when near end voice is detected so that the convergence gain μ is equal to zero and the filter coefficients are not updated. Otherwise α is set to a constant of less than one and preferably in the range of 0.8-0.95. One of skill in the art will 30 appreciate that the adaptive filter may be implemented in a variety of ways, including fixed point rather than the described floating point realization. Accordingly, the described exemplary adaptation logic is by way of example only and not by way of 35 limitation.

1

5

10

FIG. 7 shows a simplified schematic diagram of an exemplary stereo echo cancellation system that provides acoustic and electric echo cancellation. In this embodiment common reference numerals may be used to represent like components. The acoustic echo may result from a communication device such as a hands free speaker phone having a near end microphone 330 exposed to a loudspeaker 329 driven by a far end signal or a secondary audio signal. Acoustic feedback problems may also occur when the near end microphone 330 picks up the far end loudspeaker signal or a second audio signal and redirects it back to the remote end. As a result, the remote party may hear a delayed version of their own speech or of the secondary audio signal.

15

20

In the described exemplary echo cancellation system, combiner 301 combines a downsampled secondary audio signal 340(a) and far end reference signal 300 into a single 8 kHz combined reference signal 301(a). The combined reference signal 301(a) may then be input into a single 8 kHz adaptive filter 200 rather than using two adaptive filters, one at 8 kHz and the other at 48 kHz as described in connection with FIG. 5. Therefore, the described exemplary embodiment requires less memory and processing resources.

25

The described exemplary adaptive filter 200 models the impulse response of the acoustic echo path, such as for example, between the loudspeaker 329 and microphone 330 in parallel with the impulse response between the transmit channel and the receive channel of the network interface. As previously described, the adaptive filter 200 may be a linear transversal filter or other suitable finite impulse response filter.

30

35

The adaptive filter 200 converges or adapts only in the absence of near end speech. Therefore, near end speech and/or noise present on a decimated near end signal 336(a), typically referred to as the double talk condition, may cause the adaptive filter 200 to diverge. To avoid divergence, double talk logic 308 selectively enables the adaptive filter 200. The described

1
exemplary embodiment is not limited to a particular algorithm for
detecting the double talk condition. However, the alternate
5 stereo echo canceller may also utilize a sample based double talk
detection algorithm to provide rapid detection of near end speech
in the presence of far end speech. An exemplary sample based
algorithm again includes a look ahead capability so that double
talk logic may halt filter adaptation (i.e. freeze the filter
10 taps or coefficients) before the near end speech samples reaches
the difference operator 302.

15
In an exemplary double talk detection system, a downsampler
304 decimates a 96 kHz near end Tx data signal 306 at a ratio of
12:1. The downsampler 304 forwards decimated near end data
samples 304(a) to double talk logic 308 at an 8 kHz sample rate.
20 To minimize delay, the downsampler 304 does not low pass filter
the near end Tx data samples 306 prior to decimation. Aliasing
components that may be created are insignificant in that the
output of the downsampler 304(a) simply drives the double talk
detection logic 308 and is not transmitted to the far end. An
energy estimator 310 estimates the background noise level of the
decimated near end signal 304(a) and forwards the estimated level
to the double talk logic 308. The energy estimator 310 may be
a low pass filter with a long time constant, on the order of
about 10 seconds. With a long time constant the energy estimator
25 tends to track the minimum energy level of the decimated near end
signal 304(a). A second energy estimator 312 estimates the short
term energy of combined reference signal 301(a).

30 Referring to FIG. 8, in the described exemplary embodiment,
the adaptive filter models in parallel the transfer functions of
each of the three possible echo paths. Specifically, the
adaptive filter models the transfer function 400 of the
electrical echo path resulting from reflections of the far end
voice created by the impedance mismatch at the two-four wire
conversion in the hybrid. The adaptive filter also models the
35 transfer function of the acoustic echo path 402 between the

secondary audio signal being broadcast by speaker 328 and the microphone 330 as well as the transfer function of the acoustic echo path 404 between the speaker 329 broadcasting the far end voice signal and the microphone 330.

Referring back to FIG. 7, the double talk logic 308 utilizes a sample based algorithm to detect the presence of near end speech without incurring the delays associated with conventional systems in accordance with the following equation:

$$|Near| > MAX\{H_{a_v}, H_{hy_v}\} * Peak\{|V(n)|\} + H_{a_m} * Peak\{|M(n)|\} + Background_Noise(n)$$

The double talk logic 308 declares near end speech active when the absolute value of the near end Tx data signal 306 is greater than the weighted sum of the voice signal $V(n)$, secondary audio signal $M(n)$, and the background noise. In the described exemplary embodiment, a peak statistic of the voice signal $V(n)$ 327 is multiplied by the maximum of either the acoustic gain of the acoustic voice echo path from the speaker 329 to the microphone 330, or the gain of the electrical echo path caused by reflections at the hybrid.

In general, audio signals will attenuate rapidly as a function of distance from the speaker, typically on the order of about 6 db/ft. In most applications, the microphone is physically isolated from the unit's speakers to maximize this attenuation. The hybrid gain on the other hand is generally a constant typically less than about one-half. Therefore, in most applications the hybrid gain will dominate for weighing the far end voice samples.

Similarly a peak statistic of the secondary audio signal $M(n)$ is multiplied by the acoustic gain of the acoustic echo path from the second speaker 328 to the microphone 330. The background noise for a typical voice channel is on the order of about -70 dBm which is far less than average active speech

1 levels, which are typically in the order of about -25 dBm. The background noise estimate is therefore initialized to a value of about -70 dBm and thereafter periodically updated by the energy estimator 310. The peak statistic of the voice $V(n)$ and the secondary audio $M(n)$ signals are defined by the following logic:

If $\max\{A * [|V(n)|, \dots, |V(n-L+1)|]\} > V(n-1)$ then

Peak(n) = $\max\{A [|V(n)|, \dots, |V(n-L+1)|]\}$

else

Peak(n) = $d * \text{Peak}(n-1)$;

where A is a weighting function that is greater than zero and less than or equal to one. The parameter L is the number samples over which the maximum is determined, typically in the range of zero to one hundred and twenty eight samples. In the described exemplary embodiment L is on the order of about sixty four samples. The parameter d is generally a constant that is also greater than zero and less than or equal to one and in the described exemplary embodiment on the order of about 0.99. Therefore, to determine the peak statistic of the far end, the double talk logic applies a weighting factor A to the absolute value of the current sample ($V(n)$, $M(n)$) and previous L samples ($V(n-L)$, $M(n-L)$). If the maximum product is greater than the previous peak statistic $\text{Peak}(n-1)$ then the current peak statistic $\text{Peak}(n)$ is set at the maximum of the product of the weighting factor and far end samples. Otherwise the current peak statistic $\text{Peak}(n)$ is set equal to d times the value of the previous peak statistic $\text{Peak}(n-1)$.

In the described exemplary embodiment, A , L and d are empirically determined to optimize performance and computational load for a given application. For example, double logic 212 can more accurately detect near end speech if the maximum is determined over a larger number of samples L . However, computational intensity also increases with increasing number of samples L . A and d can be inversely related to the number of

1

sample L, so that A and d are smaller for larger number of samples and vice versa.

5

10

15

In the described exemplary embodiment, there is a delay associated with decimator 336. However, downsampler 304 does not low pass filter the near end TX data signal 306 and therefore does not have a corresponding delay. Therefore, the double talk logic 308 receives downsampled near end Tx data samples 304(a) with negligible delay and can process near end Tx data samples prior to their arrival at difference operator 302. Thus, the delay associated with the second decimator 336 provides a look-ahead of N samples allowing the double talk logic 308 to disable adaptation of the adaptive filter 200 N samples before the near-end signal reaches the difference operator 302. The look ahead capability N is equivalent to the delay associated with the second decimator 336 and is typically two to three 8kHz samples for a ITU-T G712 compliant system.

20

25

30

35

An interpolater 320 upsamples the far end signal from 8 to 96 kHz. The interpolater 320 may be implemented with a FIR filter that may be either minimum phase or linear phase. A digital to analog converter 322 converts the digital interpolated signal 320(a) to an analog signal 327 that is then transmitted to the near end. In one embodiment, an audio/music signal may be present on the far end signal resulting in near end acoustic feedback. The audio signal may result for example, from an integrated telephony-ready cable-set-top-box. The music signal 324 may be the result of a movie or a video game 326 playing out speaker 328 transmitted via a microphone 330 during a simultaneous two-way phone-call. Thus, the music signal 324 may act either as a noise source or as the near-end talker's voice. Similarly, an acoustic echo may result if near end microphone 330 is exposed to the output of the loudspeaker 329 when the loudspeaker is driven by a far end signal. These additional signals may create a double-talk condition at substantially all times.

1

Analog near end samples 332 having audio/music feedback and far end speech echo are received by an analog to digital converter 334 that converts the received signal 332 into a digital format. A decimator 336 downsamples the received digital signal 306 from 96 kHz to a 8 kHz rate. The decimator 336 can be implemented with a FIR filter. The difference operator 302 cancels unwanted echo by subtracting the decimated signal 336(a) from the filter output signal 200(a).

10

A decimator 340 may decimate the music signal 324 so that the sample rate of the music signal matches the sample rate of the far end reference signal 300, i.e. 8 kHz in the described exemplary embodiment. Combiner 301 combines decimated music signal 340(a) and the far end reference signal 300 into a single 8 kHz input 301(a) to the adaptive filter 200.

15

Referring to FIG. 9, an exemplary combiner 301 may again include a first buffer 350 that receives the far end reference signal 300 and a second buffer 352 that receives the decimated music signal 340(a). Buffer outputs 350(a) and 352(a) are combined in a third buffer 354 whose output 354(a) is forwarded to the adaptive filter 200. One of skill in the art will appreciate that a single buffer could be used to directly combine the decimated music signal 340(a) and far end reference signal 300.

20

25

The adaptive filter may be a linear transversal filter or other suitable finite impulse response filter. The adaptive filter 200 may again be based upon a normalized least mean square algorithm (NLMS) as described in S. Haykin, Adaptive Filter Theory, and T. Parsons, Voice and Speech Processing. The adaptive filter 200 may again use an error signal feedback control as previously described with regards to FIG. 5.

30

35

While only a number of exemplary embodiments have been described, a variety of modifications and adaptations can be made without departing from the spirit and scope of the present invention. In addition, those skilled in the art will understand

1

that various modifications may be made to the described embodiment. Moreover, to those skilled in the various arts, the invention itself herein will suggest solutions to other tasks and adaptations for other applications. It is therefore desired that the present embodiments be considered in all respects as illustrative and not restrictive, reference being made to the appended claims rather than the foregoing description to indicate the scope of the invention.

10

15

20

25

30

35